# Grade of Membership Analysis: One Possible Approach to Foundations*

Mikhail Kovtun Igor Akushevich Kenneth G. Manton
Center for Demographic Studies, Duke University, Durham, NC

H. Dennis Tolley
Department of Statistics, Brigham Young University, Provo, UT

February 1, 2008

**Abstract**

Grade of membership (GoM) analysis was introduced in 1974 [7] as a means of analyzing multivariate categorical data. Since then, it has been successfully applied to many problems. The primary goal of GoM analysis is to derive properties of individuals based on results of multivariate measurements; such properties are given in the form of the expectations of a hidden random variable (state of an individual) conditional on the result of observations.

In this article, we present a new perspective for the GoM model, based on considering distribution laws of observed random variables as realizations of another random variable. It happens that some moments of this new random variable are directly estimable from observations. Our approach allows us to establish a number of important relations between estimable moments and values of interest, which, in turn, provides a basis for a new numerical procedure.

**Keywords:** *Grade of membership analysis, latent structure analysis, multivariate categorical data, linear regression, multidimensional distribution.*

**AMS 2000 subject classifications:** *Primary 62H12; secondary 62J99.*

## 1 Introduction

The grade of membership (GoM) analysis was initially introduced in [7]; the term "Grade of Membership" is due to this article.

GoM considers $J$ discrete measurements on each individual, represented by random variables $X_1, \ldots, X_J$, with the set of outcomes of $j^{\text{th}}$ measurement being $\{1, \ldots, L_j\}$.

*This research was supported by grants from National Institute of Aging.

The goal of GoM analysis is to derive some properties of an individual based on results of measurements. We refer to this (general and informal) specification of goals as *the General GoM Problem* (GGP.) As GGP is a general concept, there may be many different but reasonable answers to the problem. The present article proposes one possible approach to GGP, which leads to notable theoretical results and allows construction of a novel numerical procedure.

Having GGP as its primary goal, GoM differs from many other statistical methods, whose goal is to discover some properties of a population. For example, in estimation of voting results the most interesting fact is how many people will vote for, or against (a candidate or an issue), and it does not matter how a particular individual votes. In contrast, in making a medical diagnosis the health of a particular individual is of interest, and it does not matter (for a particular diagnosis) how prevalent a particular health state is in a population.

Mathematically, one possibility to express GGP is to assume that there exists a hidden continuous random variable $G$ representing knowledge about individuals derivable from observations (in the diagnostic example, it is the health state of an individual.) Now one is interested in what might be said about value of $G$ based on observed values of $X_1, \dots, X_J$. More specifically, values of interest are expectations of $G$ conditional on values of random variables $X_1, \dots, X_J$, $\mathcal{E}(G \mid X_1 = x_1, \dots, X_J = x_J)$.

Considering a continuous hidden random variables resembles latent structure analysis in general, and latent trait analysis in particular (see [1, 2, 5].) The connection between GoM and latent structure analysis was mentioned in the literature ([4]; see more details in [3].) We prefer to keep the name "grade of membership analysis" because: (a) its primary goal differs from that of the latent structure analysis, and (b) GoM uses a proprietary technique and is based on special facts that are not used in latent structure analysis. However, we believe that techniques developed in the present article and results obtained here might benefit the development of latent structure analysis.

The main result of the present article, contained in section 7, is that the values of interest (i.e. conditional expectations and conditional variances) are solution of system (36), and that under modest conditions, *only* values of interest are solutions of this system. Furthermore, as corollary 7.6 shows, the system (36) can be solved by two-step process, every step of which consists of solving problem of linear algebra.

GoM analysis (as well as many flavors of latent structure analysis) employs an assumption that the problem under consideration has lower dimensionality than observed data. Our theorem 7.3 and its corollary gives a way to estimate this dimensionality directly (which usually presents a substantual problem is such kind of analysis.)

An additional advantage of our approach is that it not only establishes a way to estimate values of interest, but also provides a ground for evaluation of confidence intervals (not addressed in the present article.)

The rest of the article is organized as follows.

In section 3 we mathematically formulate the problem and define related notions. The central idea here (which is crucial for further results) is to consider

individual distribution laws as realizations of another random variable, $\beta$. We show that initial data are sufficient to estimate a set of mixed moments of this distribution up to order $J$ (the number of measurements.)

In section 4 we consider the GoM problem as a problem of finding a low-dimensional distribution and obtain basic corollaries of this hypothesis.

In section 5 we consider a hypothesis that there exists a linear regression of observed random variables $X_j$ on hidden random variable $G$. We show that this hypothesis is essentially equivalent to the one considered in previous section.

In section 6 we establish relations between distributions and moments of $\beta$ and $G$, and find transformation laws for changing their basis. The main result of this section is equation (29).

In section 7 we consider a system of equations (36). We show that values of interest are always solutions of this system, and we establish sufficient conditions under which the system (36) has *only* such solutions.

In section 8 we outline a numerical procedure for estimating values of interest and discuss its properties.

# 2 Preliminaries

## 2.1 Notation

$\mathbb{Z}$ is the set of integers, and $\mathbb{R}$ is the set of reals. $\mathbb{Z}^+$ and $\mathbb{R}^+$ are subsets of positive, and $\mathbb{Z}^{+0}$ and $\mathbb{R}^{+0}$ are subsets of nonnegative, integers and reals, respectively.

For $m, n \in \mathbb{Z}$, $[m..n]$ denotes the set of integers between $m$ and $n$: $[m..n] = \{z \in \mathbb{Z} \mid m \leq z \leq n\}$. If $m > n$, $[m..n] = \varnothing$.

$\mathbb{R}^n$ is $n$-dimensional linear space over reals, and $\mathbb{S}^n$ is a $(n-1)$-dimensional unit simplex in $\mathbb{R}^n$, $\mathbb{S}^n = \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ and } \sum_i x_i = 1\}$.

For a linear subspace $Q \subseteq \mathbb{R}^n$, $\dim(Q)$ denotes its dimension.

For $x^1, \ldots, x^p \in \mathbb{R}^n$, $\mathrm{Lin}(x^1, \ldots, x^p)$ denotes a linear subspace of $\mathbb{R}^n$ spanned by $x^1, \ldots, x^p$, and $\mathrm{rank}(x^1, \ldots, x^p)$ denotes a rank of system of vectors $x^1, \ldots, x^p$ (thus, $\mathrm{rank}(x^1, \ldots, x^p) = \dim(\mathrm{Lin}(x^1, \ldots, x^p))$.)

For $\alpha \in \mathbb{R}$ (or $\alpha \in \mathbb{Z}$) and $i \in [1..n]$, $\boldsymbol{\alpha}_i$ denotes a vector from $\mathbb{R}^n$ ($\mathbb{Z}^n$, respectively) with $i^{\text{th}}$ component equal $\alpha$ and all other components equal 0. Dimensionality of $\boldsymbol{\alpha}_i$ will be clear from context.

## 2.2 Support of measures

We consider only probabilistic measures defined on $\sigma$-algebra of Borel sets of $\mathbb{R}^n$; a measure $\mu$ is a probabilistic measure, if $\mu(\mathbb{R}^n) = 1$.

A *support* of measure $\mu$ is a closed set $A \subseteq \mathbb{R}^n$ such that $\mu(A) = 1$. We do not require a minimality of a support: if $A$ is a support of $\mu$ and $A \subseteq A'$, $A'$ is closed, then $A'$ also is a support of $\mu$.

We use $\mathrm{Supp}(\mu)$ to denote the set of all supports of $\mu$. Thus, $A \in \mathrm{Supp}(\mu)$ means "$A$ is a support of $\mu$." Note that $A \in \mathrm{Supp}(\mu)$ implies that $A$ is closed.

### 2.3 Indexing contingency tables and related objects

We need a way for indexing cells in a contingency table and for other objects having similar structure.

A contingency table for a set of $J$ discrete measurements, with $L_j$ possible outcomes for measurement $j$, is a $J$-dimensional table having $L_j + 1$ cells in dimension $j$. Index for $j^{\text{th}}$ dimension ranges from 0 to $L_j$.

More formally, let $\mathcal{L}_\infty = \{(l_1, \ldots, l_J) \mid l_j \in \mathbb{Z}^+\}$ and $\mathcal{L}^0_\infty = \{(l_1, \ldots, l_J) \mid l_j \in \mathbb{Z}^{+0}\}$, i.e. sets of $J$-dimensional vectors with positive and, respectively, nonnegative integer components. There is a one-to-one correspondence between sets of $J$ discrete measurements and vectors in $\mathcal{L}_\infty$: a vector $L = (L_1, \ldots, L_J)$ describes a set of $J$ measurements, in which measurement $j$ has $L_j$ outcomes.

For every $L \in \mathcal{L}_\infty$, let $\mathcal{L}_L = \{\ell \in \mathcal{L}_\infty \mid \ell_j \leq L_j\}$ and $\mathcal{L}^0_L = \{\ell \in \mathcal{L}^0_\infty \mid \ell_j \leq L_j\}$. If $L$ defines a set of measurements, $\mathcal{L}_L$ is a set of all possible outcomes of these measurements, and $\mathcal{L}^0_L$ is a complete set of indices for the contingency table. In addition, for every $\mathcal{J} \subseteq [1..J]$, let $\mathcal{L}^{[\mathcal{J}]}_L = \{\ell \in \mathcal{L}^0_L \mid \ell_j = 0 \iff j \in \mathcal{J}\}$. The set $\mathcal{J}$ indicates measurements that we exclude from consideration, and vector $\ell \in \mathcal{L}^{[\mathcal{J}]}_L$ contains results of all measurements except those listed in $\mathcal{J}$. Note that $\mathcal{L}^{[\varnothing]}_L = \mathcal{L}_L$ and $\mathcal{L}^0_L = \cup_{\mathcal{J} \subseteq [1..J]} \mathcal{L}^{[\mathcal{J}]}_L$.

Vector $\ell' \in \mathcal{L}^{[\mathcal{J}]}_L$ may be considered as describing a family of outcomes $\{\ell \in \mathcal{L}_L \mid \ell_j = \ell'_j$ for $j \notin \mathcal{J}\}$. Abusing notation, we will also use $\ell'$ to denote this family. More generally, we write $\ell' \in \ell''$ for $\ell' \in \mathcal{L}^{[\mathcal{J}']}_L$ and $\ell'' \in \mathcal{L}^{[\mathcal{J}'']}_L$ whenever $\ell'_j = \ell''_j$ for all $j \notin \mathcal{J}''$ (note $\ell' \in \ell''$ is possible only when $\mathcal{J}' \subseteq \mathcal{J}''$.) For $\ell \in \mathcal{L}^{[\mathcal{J}]}_L$, let $\ell^{[\mathcal{J}']}$ be a vector from $\mathcal{L}^{[\mathcal{J} \cup \mathcal{J}']}_L$ such that $\ell^{[\mathcal{J}']}_j = \ell_j$ for all $j \notin \mathcal{J}'$. We always have $\ell \in \ell^{[\mathcal{J}']}$. We write $\ell^{[j]}$, $\ell^{[j_1,j_2]}$, etc. instead of $\ell^{[\{j\}]}$, $\ell^{[\{j_1,j_2\}]}$, etc., respectively.

Let also set $|L| = \sum_j L_j$ and $|L^*| = \prod_j L_j$.

We always assume that the set of our measurements is described by a vector $L$. We drop index $L$ in notations $\mathcal{L}_L$ and $\mathcal{L}^{[\mathcal{J}]}_L$ if it does not create an ambiguity.

A contingency table may be constructed for any sample by putting in the cell with index $\ell$ the number of individuals who (a) have outcome $\ell_j$ for measurement $j$ if $\ell_j \neq 0$; and (b) have arbitrary outcomes for all other measurements. Let $N_\ell$ be a value in $\ell^{\text{th}}$ cell of contingency table. The usual summation rule for contingency tables in our notation is: for any $\mathcal{J}' \subseteq \mathcal{J} \subseteq [1..J]$, $\ell \in \mathcal{J}$, $N_\ell = \sum_{\ell' \in \mathcal{J} : \ell' \in \ell} N_{\ell'}$. Note that $N = N_{(0,\ldots,0)}$ is the sample size.

A frequency table is obtained from a contingency table by dividing the value in each cell by $N$. We use $f_\ell$ to denote a value of $\ell^{\text{th}}$ cell of frequency table. The above summation rule is applicable to frequency tables as well.

## 3 The Problem

We consider a population of a potentially infinite number of individuals, every individual being subject to $J$ measurements with discrete outcomes. With-

out loss of generality, we may assume that outcomes of $j^{\text{th}}$ measurement are $\{1, \ldots, L_j\}$.

The results of measurements on individual $i$ is a random vector $X^i = (X_1^i, \ldots, X_J^i)$, which takes values in $\mathcal{L}_L$. Such a random vector is described by a $|L|$-dimensional vector of probabilities $\beta^i = (\beta_{jl}^i)_{jl}$ ($j \in [1..J]$, and for every $j$, $l \in [1..L_j]$), where $\beta_{jl}^i = \Pr(X_j^i = l)$.

These vectors of probabilities $\beta^i$ may themselves be considered as realizations of a random vector $\beta$, with a distribution described by probabilistic measure $\mu_\beta$ on $\mathbb{R}^{|L|}$.

We start with elementary properties, which may be directly derived from definitions.

As $\beta_{jl}$ are probabilities, they satisfy

$$\text{(a)} \quad \beta_{jl} \geq 0 \qquad\qquad \text{(b)} \quad \text{for all } j\text{:} \quad \sum_{l=1}^{L_j} \beta_{jl} = 1 \tag{1}$$

Thus, a product of simplices $\mathbb{S}^L = \prod_j \mathbb{S}^{L_j} \subseteq \mathbb{R}^{|L|}$ is a support of the measure $\mu_\beta$, $\mathbb{S}^L \in \mathrm{Supp}(\mu_\beta)$.

Together with random vectors $X^i$, we consider a "composite" random vector $X = (X_1, \ldots, X_J)$: on the first step, one randomly selects a vector of probabilities $\beta$ (in accordance with measure $\mu_\beta$), and on the second step, one randomly selects outcomes in accordance with (selected on the first step) probabilities $\beta$.

According to our definitions, the conditional probability for $X_j$ is:

$$\Pr\left(X_j = l \mid \beta\right) = \beta_{jl} \tag{2}$$

from which one obtains by the law of total probability

$$\Pr\left(X_j = l\right) = \int \Pr\left(X_j = l \mid \beta\right)\, \mu_\beta(d\beta) = \int \beta_{jl}\, \mu_\beta(d\beta) \tag{3}$$

We need more assumptions about $\mu_\beta$ to derive useful properties of the model. One reasonable assumption is "local independence":

(G1) Conditional on value of $\beta$, random variables $X_1, \ldots, X_J$ are mutually independent, i.e. for every $\ell \in \mathcal{L}^0$

$$\Pr\left( \bigwedge_{j\,:\,\ell_j \neq 0} X_j = \ell_j \;\middle|\; \beta \right) = \prod_{j\,:\,\ell_j \neq 0} \Pr\left(X_j = \ell_j \mid \beta\right) \tag{4}$$

A motivation for such assumption is that all "randomness" in $X_1^i, \ldots, X_J^i$ comes from errors in measurements, and error in one measurement does not depend on error in another one. Further, "conditional on value of parameters" means that we are considering a group of individuals having the same values $\beta$; thus, every individual in a group has the same vector of probabilities $\beta$, and restriction of our random vector $X$ to this group has the vector of probabilities

$\beta$ as well; as we assumed that for every individual random variables describing him are independent, this should be true for a group of identical (with respect to our random variables) individuals. It is also wise to mention that the local independence assumption is used in almost all variations of latent structure analysis.

With the independence assumption (G1), (3) may be strengthened to:

$$\forall \ell \in \mathcal{L}^0 \quad : \quad \Pr\left( \bigwedge_{j\,:\,\ell_j \neq 0} X_j = \ell_j \right) = \int \left( \prod_{j\,:\,\ell_j \neq 0} \beta_{j\ell_j} \right) \mu_\beta(d\beta) \tag{5}$$

For every $\ell \in \mathcal{L}^0$, let the $\ell$-moment of distribution $\mu_\beta$ be

$$M_\ell(\mu_\beta) = \int \left( \prod_{j\,:\,\ell_j \neq 0} \beta_{j\ell_j} \right) \mu_\beta(d\beta) \tag{6}$$

In particular, we have $\mathcal{L}^{[1,\ldots,J]} = \{(0,\ldots,0)\}$, and $M_{(0,\ldots,0)}(\mu_\beta) = \int \mu_\beta(d\beta) = 1$.

Comparing (6) with (5), we see that the $\ell$-moment of distribution $\mu_\beta$ is equal to the probability of set of outcomes $\ell$.

For $\ell \in \mathcal{L}^{[\mathcal{J}]}$, $M_\ell(\mu_\beta)$ is $J - |\mathcal{J}|$ order mixed moment of $\mu_\beta$. The set of $\ell$-moments for all $\ell \in \mathcal{L}^0$ does not exhaust, however, the set of all moments of order up to $J$ (for example, a moment $\int \beta_{11}\beta_{12}\,\mu_\beta(d\beta)$ is not an $\ell$-moment.) At the end of the section 7 we shall discuss in more detail whether $\{M_\ell(\mu_\beta)\}_\ell$ can determine all moments of order up to $J$.

Basic statistical fact is that frequencies $f_\ell$ are consistent and efficient estimators for $M_\ell(\mu_\beta)$.

The following proposition and its corollary is an equivalent of the summation rule for contingency and frequency tables.

**Proposition 3.1** *Let $\mathcal{J}' \subseteq \mathcal{J}'' \subseteq [1..J]$. Then for every $\ell'' \in \mathcal{L}^{[\mathcal{J}'']}$*

$$M_{\ell''}(\mu_\beta) = \sum_{\ell' \in \mathcal{L}^{[\mathcal{J}']}\,:\,\ell' \in \ell''} M_{\ell'}(\mu_\beta)$$

*Proof.* For every $j_0 \in \mathcal{J}'' \setminus \mathcal{J}'$ and for every $\ell' \in \mathcal{L}^{[\mathcal{J}']}$ we have:

$$\sum_{\ell \in \mathcal{L}^{[\mathcal{J}']}\,:\,\ell \in \ell'^{[j_0]}} M_\ell(\mu_\beta) = \sum_{\ell \in \mathcal{L}^{[\mathcal{J}']}\,:\,\ell \in \ell'^{[j_0]}} \int \prod_{j \notin \mathcal{J}'} \beta_{j\ell_j}\,\mu_\beta(d\beta) =$$

$$\sum_{\ell \in \mathcal{L}^{[\mathcal{J}']}\,:\,\ell \in \ell'^{[j_0]}} \int \beta_{j_0\ell_{j_0}} \cdot \prod_{j \notin \mathcal{J}' \cup \{j_0\}} \beta_{j\ell_j}\,\mu_\beta(d\beta) =$$

$$\sum_{l=1}^{L_{j_0}} \int \beta_{j_0 l} \cdot \prod_{j \notin \mathcal{J}' \cup \{j_0\}} \beta_{j\ell_j}\,\mu_\beta(d\beta) =$$

$$\int \left( \sum_{l=1}^{L_{j_0}} \beta_{j_0 l} \right) \cdot \prod_{j \notin \mathcal{J}' \cup \{j_0\}} \beta_{j\ell_j} \, \mu_\beta(d\beta) = \\ \int 1 \cdot \prod_{j \notin \mathcal{J}' \cup \{j_0\}} \beta_{j\ell_j} \, \mu_\beta(d\beta) = M_{\ell'[j_0]}(\mu_\beta)$$

The rest of the proof is induction over the size of $\mathcal{J}'' \setminus \mathcal{J}'$. ■

**Corollary 3.2** *For every* $\mathcal{J} \subseteq [1..J]$, $\sum_{\ell \in \mathcal{L}^{[\mathcal{J}]}} M_\ell(\mu_\beta) = 1$. *In particular,* $\sum_{\ell \in \mathcal{L}} M_\ell(\mu_\beta) = 1$.

Below we consider another two (essentially equivalent) assumptions. The first one is that a support of $\mu_\beta$ is restricted to $(K-1)$-dimensional affine plane in $\mathbb{R}^{|L|}$. The second assumption is that there exists a random variable $G$ taking values in $\mathbb{R}^K$ such that there exist a linear regression of random variables $X_1, \ldots, X_J$ on $G$.

# 4 Low-dimensional distributions

The second assumption that we consider is:

(G2′) The support of $\mu_\beta$ is a $K$-dimensional linear subspace $Q$ of $\mathbb{R}^{|L|}$, and any proper subspace of $Q$ does not support $\mu_\beta$.

We include the second clause (no proper subspace of $Q$ supports $\mu_\beta$) to avoid degenerate cases. Any degenerate case may be considered as nondegenerate case for some $K' < K$.

As $\mathbb{S}^L \in \mathrm{Supp}(\mu_\beta)$, the intersection $P_\beta = Q \cap \mathbb{S}^L$ is necessarily nonempty, and this intersection supports $\mu_\beta$. In general, $P_\beta$ is $(K-1)$-dimensional polyhedral body, which has at least $K$ vertices. Let $\bar{P}_\beta$ be the $(K-1)$-dimensional affine space spanned by $P_\beta$.

Let $\Lambda = \{\lambda^1, \ldots, \lambda^K\}$ be a linear basis of $Q$. We also consider $\Lambda$ as a $|L| \times K$ matrix,

$$\Lambda = \begin{pmatrix} \lambda_{11}^1 & \ldots & \lambda_{11}^K \\ & \vdots & \\ \lambda_{JL_J}^1 & \ldots & \lambda_{JL_J}^K \end{pmatrix} \tag{7}$$

There exists considerable freedom in choosing $\Lambda$. We shall exploit it by imposing constraints on $\Lambda$. The first one is:

($\Lambda_0$) For every $k$, $\lambda^k \in \bar{P}_\beta$.

Let $g = (g_1, \ldots, g_K)$ be a vector of coordinates of a point $\beta \in Q$ in basis $\Lambda$, i.e. $\beta = \sum_{k=1}^{K} g_k \lambda^k$. Then $(\Lambda_0)$ implies

$$\sum_{k=1}^{K} g_k \lambda^k \in \bar{P}_\beta \quad \Leftrightarrow \quad \sum_{k=1}^{K} g_k = 1 \tag{8}$$

If $\Lambda$ and $\Lambda'$ are two bases of $Q$, there exists a nondegenerate $K \times K$ matrix $A = (a_k^{k'})_{k'k}$ such that $\Lambda' = \Lambda A$.

Using the fact that $(1, \ldots, 1)$ is left eigenvector of matrix $A$ corresponding to eigenvalue 1 if and only if every column of $A$ sums to 1, $\sum_k a_k^{k'} = 1$, one easily obtains the following two propositions:

**Proposition 4.1** *Let both* $\Lambda$ *and* $\Lambda'$ *satisfy* $(\Lambda_0)$. *Then* $(1, \ldots, 1)$ *is left eigenvector of matrix* $A$ *with eigenvalue* 1.

**Proposition 4.2** *Let* $\Lambda$ *satisfy* $(\Lambda_0)$ *and let* $A$ *be a nonsingular matrix with left eigenvector* $(1, \ldots, 1)$ *with eigenvalue* 1. *Then* $\Lambda' = \Lambda A$ *satisfies* $(\Lambda_0)$.

If $g$ is a coordinate vector of $\beta \in Q$ in basis $\Lambda$, $\beta = \Lambda g$, then the coordinate vector of $\beta$ in basis $\Lambda' = \Lambda A$ is $g' = A^{-1} g$.

**Remark 4.3** In matrix expressions (like $\beta = \Lambda g$ above,) we always assume that all vectors are columns. ■

Every choice of a basis $\Lambda$ induces a linear map:

$$H_\Lambda : \mathbb{R}^K \to Q, \qquad H_\Lambda(g) = \sum_k g_k \lambda^k \tag{9}$$

Note that $\Lambda$ is a matrix of linear map $H_\Lambda$ with respect to basis $\Lambda$ in $Q$ and standard unit basis in $\mathbb{R}^K$.

When the basis $\Lambda$ satisfies $(\Lambda_0)$, $H_\Lambda^{-1}(\bar{P}_\beta)$ is a unit affine plane $\bar{P}_g$ in $\mathbb{R}^K$, $\bar{P}_g = \{g \in \mathbb{R}^K \mid \sum_k g_k = 1\}$, and $P_g^\Lambda = H_\Lambda^{-1}(P_\beta)$ is a convex $(K-1)$-dimensional polyhedron in $\bar{P}_g$.

The map $H_\Lambda$ allows us to introduce a measure $\mu_g^\Lambda$ on $\bar{P}_g$, defined as:

$$\mu_g^\Lambda(B) = \mu_\beta(H_\Lambda(B)) \qquad \text{for every Borel set } B \subseteq \bar{P}_g \tag{10}$$

As $P_\beta \in \mathrm{Supp}(\mu_\beta)$, we have $P_g^\Lambda \in \mathrm{Supp}(\mu_g^\Lambda)$.

Thus, we can replace integration over $P_\beta$ by integration over $P_g^\Lambda$:

$$\int_{P_\beta} \phi(\beta) \mu_\beta(d\beta) = \int_{P_g^\Lambda} \phi(H_\Lambda(g)) \mu_g^\Lambda(dg) \tag{11}$$

for every measurable function $\phi$.

**Remark 4.4** We are trying to reflect in our notation all substantial dependencies between objects. Measure $\mu_\beta$ and polyhedron $P_\beta$, of course, do not depend on the choice of $\Lambda$; thus, no index $\Lambda$ in notation $\mu_\beta$ and $P_\beta$. On the contrary, map $H$ defined by (9) (and consequently polyhedron $P_g$ and measure $\mu_g$ defined by (10)) substantially depends on the choice of $\Lambda$ — so we use notation $H_\Lambda$, $P_g^\Lambda$, and $\mu_g^\Lambda$. However, we shall drop the index $\Lambda$ in the above notation if it is obvious from the context. ■

# 5 Linear regression hypothesis

A random variable $X_j$ has a finite range $[1..L_j]$, on which no arithmetic operations are defined. This prevents us from considering expectation, variance, etc. of $X_j$. To cope with this problem, we associate with every $X_j$ a random vector $Y_j$ taking values in $\mathbb{R}^{L_j}$ and defined as: if $X_j = l$, then $Y_j = \mathbf{1}_l$, (recall that $\mathbf{1}_l$ is a $L_j$-dimensional vector with $l^{\text{th}}$ component equals 1, and all other components equal 0.)

There is an important connection between distributions of $X_j$ and $Y_j$: if $(\beta_{jl})_l$ is a vector of probabilities of $X_j$, $\beta_{jl} = \Pr(X_j = l)$, then $\mathcal{E}(Y_j) = (\beta_{jl})_l$ (here and below $\mathcal{E}(\cdot)$ denotes expectation.) In general, for every condition $C$ we have $\mathcal{E}(Y_j \mid C) = (\Pr(X_j = l \mid C)_l$.

**Remark 5.1** As $Y_j$ is an $L_j$-dimensional vector, $\mathcal{E}(Y_j)$ is also an $L_j$-dimensional vector. We use $\mathcal{E}_m(\cdot)$ to denote $m^{\text{th}}$ component of vector expectation. ■

Thus, we have

**Proposition 5.2** *For every $j$ and for every condition $C$,* (a) $\mathcal{E}_l(Y_j \mid C) \geq 0$, *and* (b) $\sum_l \mathcal{E}_l(Y_j \mid C) = 1$.

Now we can formulate an alternative form of assumption (G2):

(G2″) There exists a random vector $G$, defined on individuals and taking values in $\mathbb{R}^K$, such that:

(a) There exists a joint distribution of $G$ and $X$.

(b) Local independence assumption holds, i.e. random variables $(X_1 \mid g)$, …, $(X_J \mid g)$ are mutually independent.

(c) For every $j$, a regression of $Y_j$ on $G$ is linear.

(d) For any $K' < K$ there is no random vector $G'$ satisfying (a)–(c).

Again, clause (d) is intended to prevent degenerate cases.

Clause (c) means that for every $j$, there exist vectors $(\lambda_{jl}^1)_l, \ldots, (\lambda_{jl}^K)_l$ such that

$$\mathcal{E}(Y_j \mid G = g) = \Big(\sum_k g_k \lambda_{jl}^k\Big)_l \tag{12}$$

or, in matrix form,

$$\mathcal{E}(Y_j \mid G=g) = \begin{pmatrix} \lambda_{j1}^1 & \dots & \lambda_{j1}^K \\ & \vdots & \\ \lambda_{jL_j}^1 & \dots & \lambda_{jL_j}^K \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ \vdots \\ g_K \end{pmatrix} = \Lambda_j \cdot g \tag{13}$$

Taking into account the relation between $\mathcal{E}(Y_j)$ and probability distribution of $X_j$, one obtains:

**Theorem 5.3** *(G2′) holds if, and only if, (G2″) holds.*

The random vector $G$, if it exists, is not defined uniquely: for every non-degenerate $K \times K$ matrix $A$, random vector $G' = A^{-1}G$ also satisfies (G2″), as:

$$\begin{aligned} \mathcal{E}(Y_j \mid G'=g') = \mathcal{E}(Y_j \mid AG'=Ag') = \mathcal{E}(Y_j \mid G=Ag') = \\ \Lambda_j \cdot (A \cdot g') = (\Lambda_j \cdot A) \cdot g' = \Lambda_j' \cdot g' \end{aligned} \tag{14}$$

This nonuniqueness corresponds to the nonuniqueness of the basis for $Q$ discussed in section 4. Again, one may choose $G$ in such a way that $(\Lambda_0)$ is satisfied.

**Corollary 5.4** *In presence of* $(\Lambda_0)$, *the possible values of* $G$ *satisfy* $\sum_k g_k = 1$. *In other words,* $G$ *takes values in a unit affine plane* $\bar{P}_g \subseteq \mathbb{R}^K$.

**Corollary 5.5** *In presence of* $(\Lambda_0)$, *a set of possible values of* $G$ *is a bounded polyhedron* $P_g \subseteq \bar{P}_g$.

We are primarily interested in what can be said about value of $G$ given outcomes of $X_1, \dots, X_J$. The most interesting values are $\mathcal{E}(G \mid X=\ell)$ and $\mathcal{D}(G \mid X=\ell)$ (were $\mathcal{D}(\cdot)$ denotes variance.) We shall derive equations for these values in the next section.

# 6 Relations between $\mu_\beta$ and $\mu_g$

As (G2′) and (G2″) are equivalent, we refer to (either of) them as (G2).

Under condition (G2) we have two distributions, $\mu_\beta$ and $\mu_g^\Lambda$, connected by (9) and (10). In this section we establish further relations between $\mu_\beta$ and $\mu_g^\Lambda$.

Throughout this section, we assume that some basis $\Lambda$ of $Q$ is fixed. We drop index $\Lambda$ in all notation; however, the reader has to keep in mind that distribution $\mu_g$, as well as all its moments, depend on $\Lambda$.

### 6.1 Unconditional moments

We can express $\ell$-moments of $\mu_\beta$ via moments of $\mu_g$. Let $\mathcal{J} \subseteq [1..J]$ and $\ell \in \mathcal{L}^{[\mathcal{J}]}$. Then:

$$
\begin{aligned}
M_\ell(\mu_\beta) = \int_{P_\beta} \left( \prod_{j \notin \mathcal{J}} \beta_{j\ell_j} \right) \mu_\beta(d\beta) = \int_{P_g} \left( \prod_{j \notin \mathcal{J}} \sum_k g_k \lambda^k_{j\ell_j} \right) \mu_g(dg) = \\
\int_{P_g} \left( \sum_{w \in \mathcal{W}^{[\mathcal{J}]}} \left( \prod_{j \notin \mathcal{J}} g_{w_j} \cdot \prod_{j \notin \mathcal{J}} \lambda^{w_j}_{j\ell_j} \right) \right) \mu_g(dg) = \\
\sum_{w \in \mathcal{W}^{[\mathcal{J}]}} \left( \left( \int_{P_g} \prod_{j \notin \mathcal{J}} g_{w_j} \, \mu_g(dg) \right) \cdot \prod_{j \notin \mathcal{J}} \lambda^{w_j}_{j\ell_j} \right) = \\
\sum_{w \in \mathcal{W}^{[\mathcal{J}]}} \left( M_w(\mu_g) \cdot \prod_{j \notin \mathcal{J}} \lambda^{w_j}_{j\ell_j} \right) \quad (15)
\end{aligned}
$$

Here $\mathcal{W}^{[\mathcal{J}]} = \{(w_1, \ldots, w_J) \mid w_j \in [1..K] \text{ if } j \notin \mathcal{J},\ w_j = 0 \text{ if } j \in \mathcal{J}\}$, and for $w \in \mathcal{W}^{[\mathcal{J}]}$,

$$
M_w(\mu_g) = \int_{P_g} \left( \prod_{j \,:\, w_j \neq 0} g_{w_j} \right) \mu_g(dg) \qquad (16)
$$

is a $(J - |\mathcal{J}|)^{\text{th}}$ order mixed moment of measure $\mu^\Lambda_g$.

Note that $\mathcal{W}^{[\mathcal{J}]} = \mathcal{L}^{[\mathcal{J}]}_{(K,\ldots,K)}$. Thus, we freely apply to $\mathcal{W}$ all notations and conventions developed for $\mathcal{L}$ in section 2.3.

The sets of indices $\mathcal{W}^{[\mathcal{J}]}$ are redundant in the sense that different elements of $\mathcal{W}^{[\mathcal{J}]}$ correspond to the same moments. However, $\mathcal{W}^{[\mathcal{J}]}$ has the following nice property:

**Proposition 6.1** *Let $\mathcal{J}' \subseteq \mathcal{J}'' \subseteq [1..J]$. Then for every $w'' \in \mathcal{W}^{[\mathcal{J}'']}$*

$$
M_{w''}(\mu_g) = \sum_{w' \in \mathcal{W}^{[\mathcal{J}']} \,:\, w' \in w''} M_{w'}(\mu_g)
$$

*Proof.* Similar to the proof of proposition 3.1. ■

**Corollary 6.2** *For every $\mathcal{J} \subseteq [1..J]$, $\sum_{w \in \mathcal{W}^{[\mathcal{J}]}} M_w(\mu_g) = 1$. In particular, $\sum_{w \in \mathcal{W}} M_w(\mu_g) = 1$.*

To handle redundancy of $\mathcal{W}$, we introduce a new set of indices.

Let $\mathcal{V}[J', K'] = \{(v_1, \ldots, v_{K'}) \mid v_k \in [0..J'] \text{ and } \sum_k v_k = J'\}$. We write $\mathcal{V}[J']$ instead of $\mathcal{V}[J', K]$ and $\mathcal{V}$ instead of $\mathcal{V}[J, K]$.

For every $\mathcal{J} \subseteq [1..J]$ and for every $v \in \mathcal{V}[J-|\mathcal{J}|]$, let $\mathcal{W}_v^{[\mathcal{J}]} = \{w \in \mathcal{W}^{[\mathcal{J}]} \mid$ for every $k$, $w$ contains exactly $v_k$ components equal $k\}$. Let also $C_v^{[\mathcal{J}]} = |\mathcal{W}_v^{[\mathcal{J}]}|$.

**Proposition 6.3**

$$\text{(a)} \quad |\mathcal{V}[J',K']| = \frac{(J'+K'-1)!}{J'!(K'-1)!} \qquad\qquad \text{(b)} \quad C_v^{[\mathcal{J}]} = \frac{(J-|\mathcal{J}|)!}{v_1!\dots v_K!}$$

*Proof.* (a) By induction over $J'+K'$ from a recurrent equality $|\mathcal{V}[J',K']| = |\mathcal{V}[J'-1,K']| + |\mathcal{V}[J',K'-1]|$.

(b) Let $J' = J - |\mathcal{J}|$. By direct computation one obtains:

$$C_v^{[\mathcal{J}]} = \binom{J'}{v_1} \cdot \binom{J'-v_1}{v_2} \dots \binom{J' - \sum_{k=1}^{K-1} v_k}{v_K}$$

from which the statement of the proposition is straightforward. ■

One corollary to proposition 6.3 is that $C_v^{[\mathcal{J}]}$ depends on $\mathcal{J}$ only through $|\mathcal{J}|$, and this value is contained in index $v$; thus, we can safely drop index $[\mathcal{J}]$ and write just $C_v$.

As for every $w, w' \in \mathcal{W}_v^{[\mathcal{J}]}$ we have $M_w(\mu_g) = M_{w'}(\mu_g)$ for every measure $\mu_g$, we can define *$v$-moments of a measure $\mu_g$* as

$$M_v(\mu_g) = M_{w_0}(\mu_g) = \int \prod_k g_k^{v_k} \mu_g(dg) \tag{17}$$

and *normalized $v$-moments of a measure $\mu_g$* as

$$\bar{M}_v(\mu_g) = \sum_{w \in \mathcal{W}_v^{[\mathcal{J}]}} M_w(\mu_g) = C_v M_{w_0}(\mu_g) = C_v \int \prod_k g_k^{v_k} \mu_g(dg) = C_v M_v(\mu_g) \tag{18}$$

In both equations, $w_0$ is an arbitrary element of $\mathcal{W}_v$. Note that both $M_v(\mu_g)$ and $\bar{M}_v(\mu_g)$ do not depend on $\mathcal{J}$.

$\mathcal{V}[J']$ is the smallest possible set of indices for $J'$-order mixed moments of $\mu_g$. Multiplier $C_v$ in (18) allows us to obtain

**Proposition 6.4** *For every $\mu_g$,*

$$\sum_{v \in \mathcal{V}[J']} \bar{M}_v(\mu_g) = \sum_{v \in \mathcal{V}[J']} C_v M_v(\mu_g) = 1$$

*Proof.* Follows from proposition 6.1. ■

Now we can continue (15):

$$\sum_{w\in\mathcal{W}^{[\mathcal{J}]}}\left(M_w(\mu_g)\cdot\prod_{j\notin\mathcal{J}}\lambda_{j\ell_j}^{w_j}\right)=\sum_{v\in\mathcal{V}[J']}\sum_{w\in\mathcal{W}_v^{[\mathcal{J}]}}\left(M_w(\mu_g)\cdot\prod_{j\notin\mathcal{J}}\lambda_{j\ell_j}^{w_j}\right)=$$
$$\sum_{v\in\mathcal{V}[J']}\left(M_v(\mu_g)\cdot\sum_{w\in\mathcal{W}_v^{[\mathcal{J}]}}\prod_{j\notin\mathcal{J}}\lambda_{j\ell_j}^{w_j}\right)=\sum_{v\in\mathcal{V}[J']}\left(M_v(\mu_g)\cdot\sum_{w\in\mathcal{W}_v^{[\mathcal{J}]}}\prod_{j\notin\mathcal{J}}\lambda_{j\ell_j}^{w_j}\right)=$$
$$\sum_{v\in\mathcal{V}[J']}\left(M_v(\mu_g^\Lambda)\cdot\Lambda(\mathcal{J},v,\ell)\right)\tag{19}$$

where $J'=J-|\mathcal{J}|$ and

$$\Lambda(\mathcal{J},v,\ell)=\sum_{w\in\mathcal{W}_v^{[\mathcal{J}]}}\prod_{j\notin\mathcal{J}}\lambda_{j\ell_j}^{w_j}\tag{20}$$

## 6.2 Conditional moments

For the joint distribution of $X=(X_1,\ldots,X_J)$ and $G$ we have, on the one hand,

$$d\Pr(G=g\wedge X=\ell)=\Pr(X=\ell\mid G=g)\cdot d\Pr(G=g)=$$
$$\left(\prod_j\sum_k g_k\lambda_{j\ell_j}^k\right)\mu_g(dg)\tag{21}$$

and, on the other hand,

$$d\Pr(G=g\wedge X=\ell)=d\Pr(G=g\mid X=\ell)\cdot\Pr(X=\ell)=$$
$$d\Pr(G=g\mid X=\ell)\cdot M_\ell(\mu_\beta)\tag{22}$$

Combining (21) and (22), one obtains

$$d\Pr(G=g\mid X=\ell)=\frac{\prod_j\sum_k g_k\lambda_{j\ell_j}^k}{M_\ell(\mu_\beta)}\,\mu_g(dg)\tag{23}$$

Similarly, for every $\mathcal{J}\subseteq[1..J]$ and for every $\ell\in\mathcal{L}^{[\mathcal{J}]}$,

$$d\Pr(G=g\mid X=\ell)=\frac{\prod_{j\notin\mathcal{J}}\sum_k g_k\lambda_{j\ell_j}^k}{M_\ell(\mu_\beta)}\,\mu_g(dg)\tag{24}$$

where for $\ell\in\mathcal{L}^{[\mathcal{J}]}$, $X=\ell$ means $\bigwedge_{j\notin\mathcal{J}}X_j=\ell_j$.

This allows us to conclude that the conditional distribution of $G\mid X=\ell$ is absolutely continuous with respect to measure $\mu_g$, and

$$p_\ell(g)=\frac{\prod_{j\notin\mathcal{J}}\sum_k g_k\lambda_{j\ell_j}^k}{M_\ell(\mu_\beta)}\tag{25}$$

is its probability density function.

Having this, we may write (for every $J'$, $v \in \mathcal{V}[J']$, $\mathcal{J} \subseteq [1..J]$, $\ell \in \mathcal{L}^{[\mathcal{J}]}$) a $v$-order mixed moment of $G$ conditional on $X = \ell$

$$\mathcal{E}(G^v \mid X = \ell) = \int g^v p_\ell(g)\, \mu_g(dg) \tag{26}$$

where $g^v$ denotes $\prod_k g_k^{v_k}$.

A special case of equation (26) for $v = (0, \ldots, 0)$ is

$$\mathcal{E}(G^{(0,\ldots,0)} \mid X^{[\mathcal{J}]} = \ell) = \int g^{(0,\ldots,0)} p_\ell(g)\, \mu_g(dg) = \int p_\ell(g)\, \mu_g(dg) = 1 \tag{27}$$

Using equation (26), we may obtain for every $j \in \mathcal{J}$ and every $l \in [1..L_j]$:

$$\begin{gathered}
\sum_k \lambda_{jl}^k \mathcal{E}(G^{v+\mathbf{1}_k} \mid X = \ell) = \sum_k \lambda_{jl}^k \int g^{v+\mathbf{1}_k} p_\ell(g)\, \mu_g(dg) = \\
\sum_k \int g^v (g_k \lambda_{jl}^k) p_\ell(g)\, \mu_g(dg) = \int g^v \left( \sum_k g_k \lambda_{jl}^k \right) \frac{\prod_{j' \notin \mathcal{J}} \sum_k g_k \lambda_{j'\ell_{j'}}^k}{M_\ell(\mu_\beta)}\, \mu_g(dg) = \\
\frac{M_{\ell+\boldsymbol{l}_j}(\mu_\beta)}{M_\ell(\mu_\beta)} \int g^v \frac{\left(\sum_k g_k \lambda_{jl}^k\right) \prod_{j' \notin \mathcal{J}} \sum_k g_k \lambda_{j'\ell_{j'}}^k}{M_{\ell+\boldsymbol{l}_j}(\mu_\beta)}\, \mu_g(dg) = \\
\frac{M_{\ell+\boldsymbol{l}_j}(\mu_\beta)}{M_\ell(\mu_\beta)} \mathcal{E}(G^v \mid X = \ell + \boldsymbol{l}_j)
\end{gathered} \tag{28}$$

By multiplying both sides of (28) by $M_\ell(\mu_\beta)$ one obtains:

$$\sum_k \lambda_{jl}^k \cdot \left( M_\ell(\mu_\beta) \cdot \mathcal{E}(G^{v+\mathbf{1}_k} \mid X = \ell) \right) = M_{\ell+\boldsymbol{l}_j}(\mu_\beta) \cdot \mathcal{E}(G^v \mid X = \ell + \boldsymbol{l}_j) \tag{29}$$

Equation (29) is the main fact that allows us to establish a numerical procedure to estimate conditional expectations. This equation holds for every $J' \geq 0$, $v \in \mathcal{V}[J']$, $\mathcal{J} \subseteq [1..J]$, $\ell \in \mathcal{L}^{[\mathcal{J}]}$, $j \in \mathcal{J}$, and $l \in [1..L_j]$.

Although equation (29) holds for every $J'$ and $\mathcal{J}$, the most important case is $J' + |\mathcal{J}| < J$: as we shall see in section 7, only conditional moments $\mathcal{E}(G^v \mid X = \ell)$, $v \in \mathcal{V}[J']$, $\ell \in \mathcal{L}^{[\mathcal{J}]}$, with $J' + |\mathcal{J}| \leq J$ may be identified from data.

## 6.3 Conditional variance

To make use of conditional expectations, one would like to know variance of $G$ conditional on outcomes of measurements. It is not hard to express variance via conditional moments:

$$\mathcal{D}_k(G \mid X=\ell) = \int (g_k - \mathcal{E}_k(G \mid X=\ell))^2 \, p_\ell(g)\, \mu_g(dg) = \\ \mathcal{E}(G^{\mathbf{2}_k} \mid X=\ell) - \mathcal{E}^2(G^{\mathbf{1}_k} \mid X=\ell) \quad (30)$$

As we shall show below, $\mathcal{E}(G^{\mathbf{2}_k} \mid X=\ell)$ can be identified only for $\ell$ having at least two components equal 0; thus, the same condition applies to identifiability of $\mathcal{D}(G \mid X=\ell)$.

## 6.4 Change of basis

Let $\Lambda' = \Lambda A$ be another basis of $Q$. Here $A$ is nonsingular $K \times K$ matrix,

$$A = \begin{pmatrix} a_1^1 & \dots & a_1^K \\ & \vdots & \\ a_K^1 & \dots & a_K^K \end{pmatrix}, \quad \text{and} \quad A^{-1} = \begin{pmatrix} \bar{a}_1^1 & \dots & \bar{a}_1^K \\ & \vdots & \\ \bar{a}_K^1 & \dots & \bar{a}_K^K \end{pmatrix} \quad (31)$$

As it was mentioned above, if a vector $\beta \in Q$ has coordinates $g$ in basis $\Lambda$, then it has coordinates $g' = A^{-1}g$ in basis $\Lambda'$. Thus, $A^{-1}$ is a matrix of transition from coordinates $g$ to coordinates $g'$. A question of interest is how the moments of $G$ are changed under this transition.

We start with moments $M_w$ for $w \in \mathcal{W}$. Let $M'_w$ be a moment calculated in coordinates $g'$. Then:

$$M'_w = M'_{(k_1,\dots,k_J)} = \int g'_{k_1} \dots g'_{k_J} \, \mu'_g(dg') = \\ \int (\bar{a}^1_{k_1} g_1 + \dots + \bar{a}^K_{k_1} g_K) \dots (\bar{a}^1_{k_J} g_1 + \dots + \bar{a}^K_{k_J} g_K)\, \mu_g(dg) = \\ \sum_{m_1=1}^{K} \dots \sum_{m_J=1}^{K} \bar{a}^{m_1}_{k_1} \dots \bar{a}^{m_J}_{k_J} \int g_{m_1} \dots g_{m_J} \, \mu_g(dg) = \\ \sum_{m_1=1}^{K} \dots \sum_{m_J=1}^{K} \bar{a}^{m_1}_{k_1} \dots \bar{a}^{m_J}_{k_J} M_{(m_1,\dots,m_J)} \quad (32)$$

which suggests that $\{M_w\}_{w\in\mathcal{W}}$ is a covariant tensor of rank $J$. Employing Einstein's convention for summation, (32) may be rewritten,

$$M'_{(k_1,\dots,k_J)} = \bar{a}^{m_1}_{k_1} \dots \bar{a}^{m_J}_{k_J} M_{(m_1,\dots,m_J)} \quad (33)$$

Tensor $\{M_w\}_{w\in\mathcal{W}}$ is symmetric, and $\{M_v\}_{v\in\mathcal{V}}$ is a set of its essential components (as for any $w \in v$, $M_w = M_v$.) Transformation rules for $M_v$ have form:

$$M'_v = \frac{1}{C_v} \sum_{v'} \left( \sum_{w\in v} \sum_{w'\in v'} \bar{a}^{w'_1}_{w_1} \dots \bar{a}^{w'_J}_{w_J} \right) M_{v'} \quad (34)$$

For the general case of conditional moments of arbitrary order, one obtains

$$\mathcal{E}(G'^v \mid X = \ell) = \frac{1}{C_v} \sum_{v'} \left( \sum_{w \in v} \sum_{w' \in v'} \bar{a}_{w_1}^{w'_1} \dots \bar{a}_{w_{J'}}^{w'_{J'}} \right) \mathcal{E}(G^{v'} \mid X = \ell) \tag{35}$$

Here $v \in \mathcal{V}[J']$ for some $J'$, $v'$ ranges over $\mathcal{V}[J']$, $w$ and $w'$ are restricted to the set $\mathcal{W}[J'] = \{(w_1, \dots, w_{J'}) \mid w_j \in [1..K]\}$, and $w \in v$ means "for every $k$, $w$ contains exactly $v_k$ components equal to $k$."

# 7 Main system of equations

Consider a system of equations,

$$\begin{cases} \sum_k \alpha_{jl}^k h_\ell^{v+\mathbf{1}_k} = h_{\ell+\boldsymbol{l}_j}^v, & J' \in [0..J-1], \quad v \in \mathcal{V}[J'], \\ & \mathcal{J} \subseteq [1..J] \,:\, |\mathcal{J}| > J', \quad \ell \in \mathcal{L}^{[\mathcal{J}]}, \\ & j \in \mathcal{J}, \quad l \in [1..L_j] \\ h_\ell^{(0,\dots,0)} = M_\ell, & \ell \in \mathcal{L}^0 \\ \sum_{v \in \mathcal{V}[J']} C_v h_{(0,\dots,0)}^v = 1, & J' \in [0..J] \end{cases} \tag{36}$$

with respect to unknowns $\alpha_{jl}^k$ and $h_\ell^v$.

Equations (29) and (27) together with proposition 6.4 give us

**Theorem 7.1** *Let $\{M_\ell\}_{\ell \in \mathcal{L}^0}$ be a set of $\ell$-moments of distribution $\mu_\beta$, which satisfies (G1) and (G2). Let also $\{\lambda^k\}_k$, $\lambda^k = (\lambda_{jl}^k)_{jl}$ be some basis of the support of $\mu_\beta$, and $\mathcal{E}(G^v \mid X = \ell)$ be conditional moments calculated with respect to this basis.*

*Then $\alpha_{jl}^k = \lambda_{jl}^k$ and $h_\ell^v = M_\ell \cdot \mathcal{E}(G^v \mid X = \ell)$ give a solution of system (36).*

In other words, all values we are interested in are solutions of system (36). Below we establish sufficient conditions for the case when (36) has *only* such solutions.

For the sake of convenience, we (abusing language) shall speak about "solution $\alpha^1, \dots, \alpha^K$", having in mind "there exist $h_\ell^v$ such that $\alpha^1, \dots, \alpha^K$ together with $h_\ell^v$ compose a solution."

Let $\alpha^1, \dots, \alpha^K$, $\alpha^k = (\alpha_{jl}^k)_{jl}$, and $h_\ell^v$ be a solution of (36). Let $\alpha'^1, \dots, \alpha'^K$ be any set of vectors such that $\mathrm{Lin}(\alpha'^1, \dots, \alpha'^K) = \mathrm{Lin}(\alpha^1, \dots, \alpha^K)$. In this case there exist a nonsingular $K \times K$ matrix $A = (a_k^{k'})_{k'k}$ such that $(\alpha'^1, \dots, \alpha'^K) = (\alpha^1, \dots, \alpha^K)A$. Let $A^{-1} = (\bar{a}_k^{k'})_{k'k}$. By straightforward computation one can show that $\alpha'^1, \dots, \alpha'^K$ together with

$$h_\ell'^v = \frac{1}{C_v} \sum_{v'} \left( \sum_{w \in v} \sum_{w' \in v'} \bar{a}_{w_1}^{w'_1} \dots \bar{a}_{w_{J'}}^{w'_{J'}} \right) h_\ell^v$$

also is a solution of (36).

Thus, we can speak about *space of solutions* $\mathrm{Lin}(\alpha^1, \dots, \alpha^K)$. Note that at this point we have no arguments for uniqueness of the space of solutions; moreover, we cannot even claim that every space of solutions have the same dimension $K$. In fact, in general case space of solutions is not unique. However, in presence of sufficient conditions that we establish below, the space of solutions is unique.

Consider equations from the first group of (36) for $v = (0, \dots, 0)$ and $\ell = (0, \dots, 0)$:

$$\Big\{ \sum_k \alpha^k_{jl} h^{\mathbf{1}_k}_{(0,\dots,0)} = h^{(0,\dots,0)}_{\boldsymbol{l}_j}, \qquad j \in [1..J], \quad l \in [1..L_j] \tag{37}$$

and substitute values for $h^{(0,\dots,0)}_{\boldsymbol{l}_j}$ from the second group of (36):

$$\Big\{ \sum_k \alpha^k_{jl} h^{\mathbf{1}_k}_{(0,\dots,0)} = M_{\boldsymbol{l}_j}, \qquad j \in [1..J], \quad l \in [1..L_j] \tag{38}$$

As $h^{\mathbf{1}_k}_{(0,\dots,0)}$ do not depend on $j$ and $l$, we obtain

**Proposition 7.2** $(M_{\boldsymbol{l}_j})_{jl} \in \mathrm{Lin}(\alpha^1, \dots, \alpha^K)$ *for every solution* $\alpha^1, \dots, \alpha^K$ *of (36).*

In other words, vector $(M_{\boldsymbol{l}_j})_{jl}$ belongs to every space of solutions.

Applying similar considerations to the case $\ell = \boldsymbol{l}'_{j'}$ for some $j' \in [1..J]$, $l' \in [1..L_{j'}]$, we obtain:

$$\Big\{ \sum_k \alpha^k_{jl} h^{\mathbf{1}_k}_{\boldsymbol{l}'_{j'}} = M_{\boldsymbol{l}'_{j'} + \boldsymbol{l}_j}, \qquad j \neq j', \quad l \in [1..L_j] \tag{39}$$

In system (39) we have equations not for all $j, l$ but only for those in which $j \neq j'$. Thus, (39) does not give us a vector from a solution space. However, it allows us to claim that for every $j'$, $l'$, a vector $(M_{\boldsymbol{l}'_{j'} + \boldsymbol{l}_j})_{jl\,:\, j \neq j'}$ (having $\sum_{j \neq j'} L_j$ components) may be extended (by adding $L_{j'}$ components) to a $|L|$-dimensional vector that belongs to $\mathrm{Lin}(\alpha^1, \dots, \alpha^K)$.

In general, for every $\ell \in \mathcal{L}^0 \setminus \mathcal{L}$ we have:

$$\Big\{ \sum_k \alpha^k_{jl} h^{\mathbf{1}_k}_{\ell} = M_{\ell + \boldsymbol{l}_j}, \qquad \ell_j = 0, \quad l \in [1..L_j] \tag{40}$$

and thus we obtain further incomplete vectors that may be completed to vectors belonging to $\mathrm{Lin}(\alpha^1, \dots, \alpha^K)$.

Let us write vector $(M_{\boldsymbol{l}_j})_{jl}$ together with incomplete vectors $(M_{\boldsymbol{l}'_{j'} + \boldsymbol{l}_j})_{jl\,:\, j \neq j'}$, etc., as columns of a matrix, with places for which we do not have moments filled by question marks. We refer to this incomplete matrix as to *moment matrix.* The moment matrix contains a column for every $\ell \in \mathcal{L}^0 \setminus \mathcal{L}$. Figure 1 gives an example of (part of) a moment matrix for the case $J = 3$, $L_1 = L_2 = L_3 = 2$. Columns in this matrix correspond to $\ell = (000)$, $(100)$, $(200)$, $(010)$, $(020)$, $(001)$, $(002)$, $(110)$; other columns are not shown.

For a moment matrix $M$ let its completion $\bar{M}$ be a matrix obtained from $M$ by replacing question marks by arbitrary numbers. The above considerations give us

$$\begin{pmatrix} M_{(100)} & ? & ? & M_{(110)} & M_{(120)} & M_{(101)} & M_{(102)} & ? & \cdots \\ M_{(200)} & ? & ? & M_{(210)} & M_{(220)} & M_{(201)} & M_{(202)} & ? & \cdots \\ M_{(010)} & M_{(110)} & M_{(210)} & ? & ? & M_{(011)} & M_{(012)} & ? & \cdots \\ M_{(020)} & M_{(120)} & M_{(220)} & ? & ? & M_{(021)} & M_{(022)} & ? & \cdots \\ M_{(001)} & M_{(101)} & M_{(201)} & M_{(011)} & M_{(021)} & ? & ? & M_{(111)} & \cdots \\ M_{(002)} & M_{(102)} & M_{(202)} & M_{(012)} & M_{(022)} & ? & ? & M_{(112)} & \cdots \end{pmatrix}$$

Figure 1: Example of moment matrix

**Theorem 7.3** *Let distribution $\mu_\beta$ satisfy (G1) and (G2). Then its moment matrix has a completion $\bar{M}$ such that* $\operatorname{rank}(\bar{M}) \le K$.

One may extend definition of rank to incomplete matrices by setting it equal to the maximal size of nonzero minor, which contains only known moments (i.e. does not contain question marks.) It is easy to see that for every completion $\bar{M}$ of $M$, inequality $\operatorname{rank}(M) \le \operatorname{rank}(\bar{M})$ holds. Thus,

**Corollary 7.4** *Let distribution $\mu_\beta$ satisfy (G1) and (G2). Then* $\operatorname{rank}(M) \le K$.

For $\mathcal{K} \subseteq \mathcal{L}^0 \setminus \mathcal{L}$, let $M[\mathcal{K}]$ denote a matrix consisting of those columns of moment matrix $M$ that correspond to elements of $\mathcal{K}$.

Now we are ready to formulate the third assumption regarding distribution $\mu_\beta$:

(G3) There exist a subset of column indices $\mathcal{K} \subseteq \mathcal{L}^0 \setminus \mathcal{L}$ such that:

- (a) For every two completions of moment matrix $\bar{M}'$ and $\bar{M}''$ satisfying $\operatorname{rank}(\bar{M}') \le K$ and $\operatorname{rank}(\bar{M}'') \le K$, the equality $\bar{M}'[\mathcal{K}] = \bar{M}''[\mathcal{K}]$ holds.
- (b) Let $\bar{M}$ be any completion of moment matrix satisfying $\operatorname{rank}(\bar{M}) \le K$. Then $\operatorname{rank}(\bar{M}[\mathcal{K}]) = K$.

Note that when (G3) holds, $\bar{M}[\mathcal{K}]$ is uniquely defined.

**Theorem 7.5** *Let distribution $\mu_\beta$ satisfy (G1), (G2), and (G3). Then for every solution of system (36)* $\operatorname{Lin}(\alpha^1, \ldots, \alpha^K) = \operatorname{Lin}(\bar{M}[\mathcal{K}])$ *(where* $\operatorname{Lin}(\bar{M}[\mathcal{K}])$ *is a linear subspace of* $\mathbb{R}^{|L|}$ *spanned by columns of* $\bar{M}[\mathcal{K}]$*.)*

*Proof.* By theorem 7.3, for every solution of (36) there exists a completion $\bar{M}'$ of $M$ such that $\operatorname{Lin}(\bar{M}') \subseteq \operatorname{Lin}(\alpha^1, \ldots, \alpha^K)$. Then $\operatorname{rank}(\bar{M}') = \dim(\operatorname{Lin}(\bar{M}')) \le \dim(\operatorname{Lin}(\alpha^1, \ldots, \alpha^K)) \le K$. Thus, by (G3), $\bar{M}'[\mathcal{K}] = \bar{M}[\mathcal{K}]$, and consequently

$\mathrm{Lin}(\bar{M}[\mathcal{K}]) \subseteq \mathrm{Lin}(\bar{M}')$. As $\dim(\mathrm{Lin}(\bar{M}[\mathcal{K}])) = \mathrm{rank}(\bar{M}[\mathcal{K}]) = K$, we obtain $\mathrm{Lin}(\alpha^1, \ldots, \alpha^K) = \mathrm{Lin}(\bar{M})$. ■

**Corollary 7.6** *Let distribution $\mu_\beta$ satisfy (G1), (G2), and (G3). Then:*

*(a) To obtain a solution of (36), it is enough to take $\alpha^1, \ldots, \alpha^K$ equal to any basis of $\bar{M}[\mathcal{K}]$ (e.g., equal to any $K$ linearly independent columns of $\bar{M}[\mathcal{K}]$.)*

*(b) Any other solution $\alpha'^1, \ldots, \alpha'^K$ is obtained from the above one by multiplying it by nonsingular $K \times K$ matrix.*

*(c) Every solution $\alpha^1, \ldots, \alpha^K$ is a basis of $Q$, a support of $\mu_\beta$.*

*Proof.* (a) and (b) are obvious.

To prove (c), consider that by theorem 7.1, every basis of $Q$ is a solution of (36). By (b), all solutions are bases of the same linear subspace of $\mathbb{R}^{|L|}$. Thus, every solution is basis of $Q$. ■

By theorem 7.5 and its corollary, assumption (G3) is sufficient to identify a support of $\mu_\beta$. It looks like it is close to a necessary condition, as in many cases where (G3) is violated, we were able to construct a different distribution $\mu'_\beta$, which has the same $\ell$-moments as $\mu_\beta$ (and therefore $\mu'_\beta$ is indistinguishable from $\mu_\beta$ based on available observations.) However, the exact formulation of necessary conditions for identifiability of support of $\mu_\beta$ is an open question.

To verify whether condition (G3) holds, it is enough to analyze the moment matrix. Numerous practical methods might be suggested to do such verification. Without going into details, we demonstrate by example one possibility.

**Example 7.7** Consider a case $J = 3$, $L_1 = L_2 = L_3 = 2$; thus, $\mathbb{R}^{|L|} = \mathbb{R}^6$. Consider a distribution $\mu_\beta$ concentrated in three points, $\beta^{(1)}$, $\beta^{(2)}$, and $\beta^{(3)}$, with every point having probability $\frac{1}{3}$ (see figure 2). As $\beta^{(3)} = \frac{1}{2}\beta^{(1)} + \frac{1}{2}\beta^{(2)}$ and $\{\beta^{(1)}, \beta^{(2)}, \beta^{(3)}\} \in \mathrm{Supp}(\mu_\beta)$, (G2) is satisfied for $K = 2$.

The moment matrix $M$ of this distribution (which corresponds to moment matrix on figure 1) is shown on figure 2.

A submatrix of $M$ consisting of rows 3 and 4 and columns 1 and 2 is nonsingular, and therefore $x$ and $y$ such that

$$\mathrm{column1} \cdot x \;+\; \mathrm{column2} \cdot y \;=\; \mathrm{column7}$$

are uniquely defined; they are $x = \frac{131}{160}$ and $y = -\frac{99}{160}$. This allows construction of the only possible completion of column 7, which is shown on figure 2.

Thus, column 1 and (completed) column 7 give a basis for a support of $\mu_\beta$. It is easy to see that $\mathrm{Lin}(\mathrm{column1}, \mathrm{column7}) = \mathrm{Lin}(\beta^{(1)}, \beta^{(2)}, \beta^{(3)})$, as one would expect.

Vectors "column1" and "completed column7" do not satisfy condition $(\Lambda_0)$. To obtain a basis satisfying $(\Lambda_0)$, one can take $\alpha^1 = \mathrm{column1}$ and $\alpha^2 = \mathrm{column7} \cdot \frac{40}{19}$. Vectors $\alpha^1$ and $\alpha^2$ are shown on figure 2.

$$\beta^{(1)} = \begin{pmatrix} \beta^{(1)}_{11} \\ \beta^{(1)}_{12} \\ \beta^{(1)}_{21} \\ \beta^{(1)}_{22} \\ \beta^{(1)}_{31} \\ \beta^{(1)}_{32} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \frac{1}{3} \\ \frac{2}{3} \\ \frac{4}{5} \\ \frac{1}{5} \end{pmatrix}, \quad \beta^{(2)} = \begin{pmatrix} \beta^{(2)}_{11} \\ \beta^{(2)}_{12} \\ \beta^{(2)}_{21} \\ \beta^{(2)}_{22} \\ \beta^{(2)}_{31} \\ \beta^{(2)}_{32} \end{pmatrix} = \begin{pmatrix} \frac{1}{9} \\ \frac{8}{9} \\ \frac{3}{5} \\ \frac{2}{5} \\ \frac{1}{4} \\ \frac{3}{4} \end{pmatrix}, \quad \beta^{(3)} = \begin{pmatrix} \beta^{(3)}_{11} \\ \beta^{(3)}_{12} \\ \beta^{(3)}_{21} \\ \beta^{(3)}_{22} \\ \beta^{(3)}_{31} \\ \beta^{(3)}_{32} \end{pmatrix} = \begin{pmatrix} \frac{5}{9} \\ \frac{4}{9} \\ \frac{7}{15} \\ \frac{8}{15} \\ \frac{21}{40} \\ \frac{19}{40} \end{pmatrix}$$

$$M = \begin{pmatrix}
\frac{5}{9} & ? & ? & \frac{89}{405} & \frac{136}{405} & \frac{403}{1080} & \frac{197}{1080} & ? & \cdots \\
\frac{4}{9} & ? & ? & \frac{20}{81} & \frac{16}{81} & \frac{41}{270} & \frac{79}{270} & ? & \cdots \\
\frac{7}{15} & \frac{89}{405} & \frac{20}{81} & ? & ? & \frac{397}{1800} & \frac{443}{1800} & ? & \cdots \\
\frac{8}{15} & \frac{136}{405} & \frac{16}{81} & ? & ? & \frac{137}{450} & \frac{103}{450} & ? & \cdots \\
\frac{21}{40} & \frac{403}{1080} & \frac{41}{270} & \frac{397}{1800} & \frac{137}{450} & ? & ?[\leftarrow \frac{191}{960}] & \frac{151}{1080} & \cdots \\
\frac{19}{40} & \frac{197}{1080} & \frac{79}{270} & \frac{443}{1800} & \frac{103}{450} & ? & ?[\leftarrow \frac{53}{192}] & \frac{259}{3240} & \cdots
\end{pmatrix}$$

$$\alpha^1 = \begin{pmatrix} \alpha^1_{11} \\ \alpha^1_{12} \\ \alpha^1_{21} \\ \alpha^1_{22} \\ \alpha^1_{31} \\ \alpha^1_{32} \end{pmatrix} = \begin{pmatrix} \frac{5}{9} \\ \frac{4}{9} \\ \frac{7}{15} \\ \frac{8}{15} \\ \frac{21}{40} \\ \frac{19}{40} \end{pmatrix}, \qquad \alpha^2 = \begin{pmatrix} \alpha^2_{11} \\ \alpha^2_{12} \\ \alpha^2_{21} \\ \alpha^2_{22} \\ \alpha^2_{31} \\ \alpha^2_{32} \end{pmatrix} = \begin{pmatrix} \frac{197}{513} \\ \frac{316}{513} \\ \frac{443}{855} \\ \frac{412}{855} \\ \frac{191}{456} \\ \frac{265}{456} \end{pmatrix}$$

Figure 2: Illustration to example 7.7

(Calculations for this and subsequent examples were done with Waterloo Maple™ v.7.00.) ■

The second question is whether $h^v_\ell$ may be uniquely determined from (36) given a solution $\alpha^1, \ldots, \alpha^K$. In general, the answer is negative: not all $h^v_\ell$ may be uniquely determined. However, a number of the most important values always may be determined uniquely, as the following theorem shows.

**Theorem 7.8** *Let $\alpha^1, \ldots, \alpha^K$ be a solution of (36), and let set of index pairs $j_1 l_1, \ldots, j_K l_K$, with $l_k \in [1..L_{j_k}]$, be chosen so that the matrix $(\alpha^{k'}_{j_k l_k})_{k'k}$ is*

*nonsingular (this is always possible as* $\mathrm{rank}(\alpha^1, \ldots, \alpha^K) = K$*.) Let* $\mathcal{J}_0 = \{j_1, \ldots, j_K\}$ *(note that* $|\mathcal{J}_0|$ *may be less than* $K$*.) Then:*

*(a) For every* $\mathcal{J}$ *such that* $\mathcal{J}_0 \subseteq \mathcal{J}$*, for every* $\ell \in \mathcal{L}^{[\mathcal{J}]}$*, and for every* $k \in [1..K]$*, the conditional expectation* $\mathcal{E}(G_k \mid X = \ell)$ *is uniquely defined.*

*(b) Let, in addition, there exist* $j_0 \notin \mathcal{J}_0$ *and* $l_0 \in [1..L_{j_0}]$ *such that every* $K \times K$ *submatrix of* $(K+1) \times K$ *matrix* $(\alpha^{k'}_{j_k l_k})_{k' \in [1..K], k \in [0..K]}$ *is nonsingular. Then for every* $\mathcal{J}$ *such that* $\mathcal{J}_0 \cup \{j_0\} \subseteq \mathcal{J}$*, for every* $\ell \in \mathcal{L}^{[\mathcal{J}]}$*, and for every* $k \in [1..K]$*, the conditional variance* $\mathcal{D}(G_k \mid X = \ell)$ *is uniquely defined.*

*Proof.* (a) Consider a subsystem of (36):

$$\left\{ \sum_{k'} \alpha^{k'}_{j_k l_k} h_\ell^{\mathbf{1}_{k'}} = M_{\ell + (\boldsymbol{l}_k)_{(j_k)}}, \qquad k = 1, \ldots, K \right.$$

By theorem 7.1, $h_\ell^{\mathbf{1}_{k'}} = M_\ell \cdot \mathcal{E}(G_k \mid X = \ell)$ is a solution of this system, and by assumption of the theorem, there are no other solutions.

(b) By part (a) of the theorem, for every $k_0 \in [1..K]$ and every $k \in [1..K]$, values $h^{\mathbf{1}_{k_0}}_{\ell + (\boldsymbol{l}_k)_{(j_k)}}$ are uniquely determined from (36). Now consider a subsystem of (36):

$$\left\{ \sum_{k'} \alpha^{k'}_{j_k l_k} h_\ell^{\mathbf{1}_{k_0} + \mathbf{1}_{k'}} = h^{\mathbf{1}_{k_0}}_{\ell + (\boldsymbol{l}_k)_{(j_k)}}, \qquad k = 1, \ldots, K; \quad k_0 = 1, \ldots, K \right.$$

By theorem 7.1, $h_\ell^{\mathbf{1}_{k_0} + \mathbf{1}_{k'}} = M_\ell \cdot \mathcal{E}(G^{\mathbf{1}_{k_0} + \mathbf{1}_{k'}} \mid X = \ell)$ is a solution of this system, and by assumption of the theorem, there are no other solutions. This is enough to calculate $\mathcal{D}(G_k \mid X = \ell)$ using formula (30). ■

**Example 7.9** We continue example 7.7. Consider a subsystem of (36):

$$\begin{cases} \alpha^1_{21} h^{(1,0)}_{(1,0,0)} + \alpha^2_{21} h^{(0,1)}_{(1,0,0)} = M_{(1,1,0)} \\ \alpha^1_{22} h^{(1,0)}_{(1,0,0)} + \alpha^2_{22} h^{(0,1)}_{(1,0,0)} = M_{(1,2,0)} \end{cases}, \quad \text{or} \quad \begin{cases} \frac{7}{15} h^{(1,0)}_{(1,0,0)} + \frac{443}{855} h^{(0,1)}_{(1,0,0)} = \frac{89}{405} \\ \frac{8}{15} h^{(1,0)}_{(1,0,0)} + \frac{412}{855} h^{(0,1)}_{(1,0,0)} = \frac{136}{405} \end{cases}$$

Solving this system gives

$$h^{(1,0)}_{(1,0,0)} = \frac{131}{99}, \qquad h^{(0,1)}_{(1,0,0)} = -\frac{76}{99}$$

and, as $h_\ell^v = M_\ell \cdot \mathcal{E}(G^v \mid X = \ell)$,

$$\mathcal{E}(G^{(1,0)} \mid X = (1,0,0)) = \frac{131}{55}, \qquad \mathcal{E}(G^{(0,1)} \mid X = (1,0,0)) = -\frac{76}{55}$$

Considering similar subsystems, one obtains, in particular,

$$h^{(1,0)}_{(1,0,1)} = \frac{3089}{2970}, \quad h^{(0,1)}_{(1,0,1)} = -\frac{2641}{3960}, \quad h^{(1,0)}_{(1,0,2)} = \frac{841}{2970}, \quad h^{(0,1)}_{(1,0,2)} = -\frac{133}{3960}$$

Substituting these values into subsystems,

$$\begin{cases} \alpha^1_{31} h^{(2,0)}_{(1,0,0)} + \alpha^2_{31} h^{(1,1)}_{(1,0,0)} = h^{(1,0)}_{(1,0,1)} \\ \alpha^1_{32} h^{(2,0)}_{(1,0,0)} + \alpha^2_{32} h^{(1,1)}_{(1,0,0)} = h^{(1,0)}_{(1,0,2)} \end{cases}, \quad \begin{cases} \alpha^1_{31} h^{(1,1)}_{(1,0,0)} + \alpha^2_{31} h^{(0,2)}_{(1,0,0)} = h^{(0,1)}_{(1,0,1)} \\ \alpha^1_{32} h^{(1,1)}_{(1,0,0)} + \alpha^2_{32} h^{(0,2)}_{(1,0,0)} = h^{(0,1)}_{(1,0,2)} \end{cases}$$

one finds,

$$h^{(2,0)}_{(1,0,0)} = \frac{3323}{726}, \qquad h^{(1,1)}_{(1,0,0)} = -\frac{7087}{2178}, \qquad h^{(0,2)}_{(1,0,0)} = \frac{1895}{726}$$

and thus,

$$\mathcal{E}(G^{(2,0)} \mid X = (1,0,0)) = \frac{9969}{1210}, \qquad \mathcal{E}(G^{(0,2)} \mid X = (1,0,0)) = \frac{1083}{242}$$

This allows us calculate conditional variances (using formula (30)):

$$\mathcal{D}(G_1 \mid X = (1,0,0)) = \frac{15523}{6050}, \qquad \mathcal{D}(G_2 \mid X = (1,0,0)) = \frac{15523}{6050}$$

Table 1 summarize conditional expectations and conditional variances that may be calculated in our example. Although all values are *exact* rational numbers, we used decimal notation to make comparison of values easier. We also put standard deviations in the table instead of variances.

As we have mentioned, there are many choices for basis for the support of distribution $\mu_\beta$. Another possibility is to take $\{\beta^{(1)}, \beta^{(2)}\}$ as a basis. The result of calculations in this basis is given in table 2. One can see that, although numbers are different, their relative position remains the same. ■

**Remark 7.10** The standard deviations in the above example are relatively large. This is direct consequence of the fact that in this example we have too small number of measurements. When number of measuremnents increases, the standard deviation becomes smaller and smaller. ■

**Remark 7.11** Theorem 7.8 guarantees that it is always possible to find $J - K$ measurements such that expectations of $G$ conditional on outcomes of these measurements may be uniquely determined from the system (36). The possibility of determining conditional variances is not guaranteed by this theorem, however. In many practical cases that we have investigated, conditions of the part (b) of theorem 7.8 are satisfied, and conditional variances can be found (as in example 7.9.) The exact conditions for determinability of conditional variances is an open question. ■

Table 1: Conditional expectations and standard deviations calculated in basis $\{\alpha^1, \alpha^2\}$.

| $\ell$ | $\mathcal{E}(G_1 \mid X = \ell)$ | $\sigma(G_1 \mid X = \ell)$ | $\mathcal{E}(G_2 \mid X = \ell)$ | $\sigma(G_2 \mid X = \ell)$ |
|---|---|---|---|---|
| $(1, 0, 0)$ | 2.3818 | 1.6018 | -1.3818 | 1.6018 |
| $(2, 0, 0)$ | -0.7273 | 1.2214 | 1.7273 | 1.2214 |
| $(0, 1, 0)$ | 0.5065 | 2.0571 | 0.4935 | 2.0571 |
| $(0, 2, 0)$ | 1.4318 | 2.0709 | -0.4318 | 2.0709 |
| $(0, 0, 1)$ | 1.9048 | 1.9122 | -0.9048 | 1.9122 |
| $(0, 0, 2)$ | 0.0000 | 1.8642 | 1.0000 | 1.8642 |

Table 2: Conditional expectations and standard deviations calculated in basis $\{\beta^{(1)}, \beta^{(2)}\}$.

| $\ell$ | $\mathcal{E}(G_1 \mid X = \ell)$ | $\sigma(G_1 \mid X = \ell)$ | $\mathcal{E}(G_2 \mid X = \ell)$ | $\sigma(G_2 \mid X = \ell)$ |
|---|---|---|---|---|
| $(1, 0, 0)$ | 0.7667 | 0.3091 | 0.2333 | 0.3091 |
| $(2, 0, 0)$ | 0.1667 | 0.2357 | 0.8333 | 0.2357 |
| $(0, 1, 0)$ | 0.4048 | 0.3970 | 0.5952 | 0.3970 |
| $(0, 2, 0)$ | 0.5833 | 0.3997 | 0.4167 | 0.3997 |
| $(0, 0, 1)$ | 0.6746 | 0.3690 | 0.3254 | 0.3690 |
| $(0, 0, 2)$ | 0.3070 | 0.3598 | 0.6930 | 0.3598 |

**Remark 7.12** By computations similar to used in (15) and (19), one obtains for every family of $J' \le J$ index pairs $j_1 l_1, \dots, j_{J'} l_{J'}$ with $l_p \in [1..L_{j_p}]$ ($j_p$ is not necessarily different from $j_{p'}$ for $p \ne p'$)

$$\int \beta_{j_1 l_1} \cdot \ldots \cdot \beta_{j_{J'} l_{J'}} \, \mu_\beta(d\beta) = \sum_{v \in \mathcal{V}[J']} M_v(\mu_g) \cdot \tilde{\Lambda}(v, j_1, l_1, \dots, j_{J'}, l_{J'})$$

where $\tilde{\Lambda}(v, j_1, l_1, \dots, j_{J'}, l_{J'})$ depends only on $\lambda^k_{jl}$. Thus, if the system (36) allows unique determination of all unknowns $h^v_\ell$, all moments of order up to $J$ of $\mu_\beta$ can be identified. This is the case, for instance, in the example 7.9.

We do not know now whether there exist some regular conditions under which the system (36) has a unique solution (modulo change of basis.) Examples that we have considered suggest that in a regular case system (36) *never* has a unique solution whenever $K > L_j$ at least for one $j$. (However, as theorem 7.8

shows, many values of interest always may be uniquely determined.) The exact description of parameters that may be uniquely identified based on system (36), and to what degree the freedom in choosing other parameters may be reduced, is a subject for further investigation. ■

# 8 Numerical procedure

We have established a number of *precise* relations between values of interest (i.e. expectation and variance of hidden random vector $G$ conditional on outcomes of measurements) and moments of (unknown) distribution $\mu_\beta$, which are directly estimable from observations. The most important of these relations are given by equations (29), and by system of equations (36). This relations suggest a numerical procedure for estimation of values of interest.

As was mentioned above, sample frequences $f_\ell$ are consistent estimators for moments $M_\ell(\mu_\beta)$. Thus, applying the least squares method to the system

$$\begin{cases} \sum_k \alpha_{jl}^k h_\ell^{v+\mathbf{1}_k} = h_{\ell+\boldsymbol{l}_j}^v, & J' \in [0..J-1],\quad v \in \mathcal{V}[J'], \\ & \mathcal{J} \subseteq [1..J] : |\mathcal{J}| > J',\quad \ell \in \mathcal{L}^{[\mathcal{J}]}, \\ & j \in \mathcal{J},\quad l \in [1..L_j] \\ h_\ell^{(0,\dots,0)} = f_\ell, & \ell \in \mathcal{L}^0 \\ \sum_{v\in\mathcal{V}[J']} C_v h_{(0,\dots,0)}^v = 1, & J' \in [0..J] \end{cases} \tag{41}$$

one obtains consistent estimators for a basis $\{\lambda^k\}_k$ and conditonal expectations of $G$.

The consistency of estimators obtained from (41) is almost straightforward corollary to consistency of estimators $f_\ell$. The rate of convergence is more delicate question (as a rate of convergence of $f_\ell$ depends on $\ell$,) and deserves separate investigation.

Theorem 7.5 suggests another, two-step way for finding solutions of (41). On the first step, one finds a basis from frequency matrix (i.e. moment matrix with frequences substituted for moments.) After basis is obtained, (41) turns to be a linear system with respect to $h_\ell^v$. This way requires significantly less computations, but its convergence properties have to be more carefully investigated.

One question regarding numerical procedure is the choice of value of $K$ for which system (41) should be solved. Theorem 7.3 and its corollary suggest that one has to take $K$ equal to the rank of the frequency matrix (modulo possible deviations of frequencies from the true moments.)

Another question is how a numerical algorithm has to deal with is nonuniqueness of basis $\{\lambda^k\}_k$. In general, there are $K^2$ degrees of freedom in choice of a basis. Imposing condition $(\Lambda_0)$ reduces this number to $K(K-1)$. One can consider additional restrictions on choice of basis:

$(\Lambda_1)$ For every $k$, unconditional expectation $\mathcal{E}_k(G)$ equals $\frac{1}{K}$.

$(\Lambda_2)$ The map $H_\Lambda$ is isometry of $\bar{P}_g$ and $\bar{P}_\beta$ (with respect to euclidean distance.)

The firts one corresponds to restricting transformations of $\bar{P}_g$, described by matrix $A$ (introduced in section 4) to those having the "center" point $(\frac{1}{K}, \ldots, \frac{1}{K})$ of $\bar{P}_g$ fixed. The second restriction guarantees that variances do not depend on the choice of basis, and variances calculated in $g$-space coincide with variances calculated in $\beta$-space.

Imposing similar restrictions based on higher order moments, one might fully eliminate nonuniqueness.

Estimation of variances is another source of problems. Formula (30) is of theoretical importance, as it demonstrates that we have enough information to estimate variances. However, it hardly can be used for numerical computations as it involves *differences* of values that we can only approximately estimate. We are working on finding a better way to estimate variances.

# 9 Conclusion

We developed a novel approach to analysis of categorical data based on considering distribution laws of observed random variables as realizations of another random variable. This starting point leads to a fruitful development.

In the present article, we were able to obtain system of equations (36) and establish its properties in theorems 7.1–7.8. This provides a base for an efficient numerical procedure that gives (one form of) an answer to General GoM Problem.

We also believe that the approach in general, and our results regarding system (36) in particular, may be successfully applied in other domains of statistics, especially in latent structure analysis.

# References

[1] Bartholomew, D.J., & Knott, M. (1999) *Latent Variable Models and Factor Analysis.* 2nd ed., London: Arnold; New York: Oxford University Press.

[2] Clogg, C.C. (1995) *Latent Class Models.* In "Handbook of Statistical Modeling for the Social and Behavioral Sciences", Arminger, G., Clogg, C.C., & Sobel, M.E., eds., New York: Plenum Press, 311–360.

[3] Erosheva, E.A. (2002) *Grade of Membership and Latent Structure Models With application to Disability Survey Data.* Ph.D. Thesis, Department of Statistics, Carnegie Mellon University. Available at `http://www.stat.cmu.edu/~fienberg/NLTCS_Models/Erosheva-thesis-2002.pdf`

[4] Haberman, S.J. (1995) *Book review of 'Statistical Applications Using Fuzzy Sets', by Kenneth G. Manton, Max A. Woodbury, and H. Dennis Tolley.* Journal of the American Statistical Association, **90** (431), 1131–1133.

[5] Heinen, T. (1996) *Latent class and discrete latent trait models: similarities and differences.* Thousand Oaks, Calif.: Sage Publications.

[6] Neyman, J., & Scott, E.L. (1948) *Consistent Estimates Based on Partially Consistent Observations.* Econometrica, **16** (1), 1-32.

[7] Woodbury, M., & Clive, J. (1974) Clinical pure types as a fuzzy partition. *Journal of Cybernetics* **4**, 111–121.

Mikhail Kovtun,
Igor Akushevich,
Kenneth G. Manton

Center for Demographic Studies
Duke University
2117 Campus Drive
Durham, NC 27708
E-mail: mkovtun@cds.duke.edu
aku@cds.duke.edu
kmg@cds.duke.edu

H. Dennis Tolley

Department of Statistics
Brigham Young University
230 TMCB
Provo, UT 84602
E-mail: tolley@byu.edu